1987-02

Robust Empirical Bayes Analyses of Event Rates

Gaver, Donald P.; O'Muircheartaigh, I.G.

American Statistical Association and the American Society for Quality Control

Gaver, D. P., & O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analyses of event rates. Technometrics, 29(1), 1-15.
http://hdl.handle.net/10945/62192

TECHNOMETRICS, FEBRUARY 1987, VOL. 29, NO. 1

Robust Empirical Bayes Analyses of Event Rates

Donald P. Gaver

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943-5100

I. G. O'Muircheartaigh

University College
Galway
Ireland

A collection of I similar items generates point event histories; for example, machines experience failures or operators make mistakes. Suppose the intervals between events are modeled as iid exponential (λ_i), or the counts as Poisson $(\lambda_i t_i)$, for the ith item. Furthermore, so as to represent between-item variability, each individual rate parameter, λ_i, is presumed drawn from a fixed (super) population with density g_λ $(\cdot\,; \boldsymbol{\theta})$, $\boldsymbol{\theta}$ being a vector parameter: a parametric empirical Bayes (PEB) setup. For g_λ, specified alternatively as log-Student $t(n)$ or gamma, we exhibit the results of numerical procedures for estimating superpopulation parameters $\boldsymbol{\theta}$ and for describing pooled estimates of the individual rates, λ_i, obtained via Bayes's formula. Three data sets are analyzed, and convenient explicit approximate formulas are furnished for λ_i estimates. In the Student-t case, the individual estimates are seen to have a robust quality.

KEY WORDS: Empirical Bayes; Robustness; Poisson process; Reliability.

1. INTRODUCTION

In reliability problems, but also in studies of logistics and congestion systems and elsewhere, it is common to encounter collections of nominally similar equipments or other entities that generate point events at similar, but not identical, rates. The questions then arise as to whether evidence for differences in the rates can be elicited from rate event data on all members of such a collection and how the data can be well used to provide strengthened estimates of the underlying true rates of the individual equipments. If all equipments seem to have about the same failure rate, then there should be little harm in calculating a simple pooled rate and quoting it for all members, whereas if evidence of considerable difference between members is present, then the individual rates seem most appropriate. Some form of compromise will be worthwhile for intermediate cases. The following general setup formalizes such situations and provides compromise estimates that tend to pool the data in a sensible manner.

Consider a collection of I equipments or other units that independently generate events in accordance with Poisson processes of constant rate λ_i. Observations of these processes are available: for unit i, s_i $(= 0, 1, 2, \ldots)$ events have been observed over an exposure time interval t_i $(i = 1, 2, \ldots, I)$. To describe the possible variability between rates, characterize λ_i as the independent realization of a random variable λ with fixed parametric density function g_λ $(\cdot\,; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a generic vector parameter. The density g_λ can be said to describe a superpopulation of rate parameters, sample values from which have been bestowed on the units of interest. The first objective of the analysis will be to use all available data to estimate the prevailing superpopulation parameters, $\boldsymbol{\theta}$; the second is to mobilize the estimated superpopulation parameters, possibly by Bayes's formula or an alternative, to provide suitably pooled or shrunken estimates for individual rates. Both point and interval estimates are desirable. Models of the aforementioned type are called *parametric empirical Bayes* (PEB) models [see Morris (1983) for a review with various references]. Our present approach emphasizes superpopulation specifications that lead to robust estimates in the sense that the possibility of widely discrepant rates or exponential parameters is automatically dealt with by the superpopulation form. Such performance can be called *discrepancy tolerant*; it resembles in various ways the behavior of modern robust location estimation and regression techniques (see Mosteller and Tukey 1977); we call our procedure *robust parametric empirical Bayes* (RPEB). General ideas of robust Bayesian analyses were described by Berger (1980, 1984); Albert (1979) considered the Poisson case. The simultaneous estimation of Poisson means has been considered by many authors; a recent high-level account is by Johnstone (1984), who provided many references. See also Martz and Waller (1982), who described work in the system reliability area.

The model described is simplistic in recognizing two sorts of variability in point event data: the ordinary, Poissonian sampling variation of observations around a given λ value ("within" variation) and the variation of the individual λ values around a fixed, unknown value ("between" variation). Many elaborations are possible. A natural possibility to consider is that rate variation is controlled in part by operational factors such as temperature, vibration, maintenance frequency and adequacy, and so forth, describable by a regression model. Another possibility is that individual rates are themselves realizations of random processes, possibly with the addition of trends, thus requiring representation of time-dependent extra-Poisson or overdispersed variations (see Burridge 1981; Cox and Lewis 1966; Gaver 1963; McCullagh and Nelder 1983, pp. 131–133; Reynolds and Savage 1971). The present article does not deal with these issues, but extensions are in progress.

The emphasis of this article is data-analytical. Algorithms are first constructed for estimation of superpopulation parameters; confidence regions associated with these are constructed and displayed graphically. The superpopulation parameter estimates are then applied to compute point and associated interval estimates of individual rate parameters. Much of this latter process is carried out numerically and displayed graphically as well. New shortcut and computationally economical approximate solutions to the preceding problems are furnished and compared to complete Bayes solutions. The procedures are applied to three sets of reliability data, and the results are discussed. Despite the formal probabilistic underpinnings described for the procedure, it seems reasonable to apply the methods in an exploratory fashion to probe for structure in data sets. This process has been briefly illustrated for one example.

2. SOME ILLUSTRATIVE DATA SETS

Here are some data sets that serve to motivate our later analyses.

2.1 Failure Rates of Air Conditioning Equipment

A classical data set to which our analysis appears applicable is that of failures of air conditioning equipment on 13 Boeing 720 aircraft; these data were originally provided by Proschan (1963) and have been much studied. We consider an initial analysis that takes each aircraft to have a constant individual mean time to failure, λ_i^{-1} ($i = 1, 2, 3, \ldots, I = 13$) and an iid exponential time to failure. The data can be summarized in terms of numbers of failures over an exposure time (see Cox and Lewis 1966; for further discussion, see Cox and Snell 1980).

Note that actual time-to-individual-failure data are available for each individual equipment. An initial data analysis of each unit's failure pattern failed to reveal substantial trend or evidence of departure from an exponential failure law. The likelihood function for λ_i, an individual exponential law parameter, is of the Poisson-gamma form with s_i the sufficient statistic, so the data are presented as such in Section 3 and provisionally analyzed to elicit between-λ_i variability. The columns headed r_i in the tables include the raw quotient (individual maximum likelihood estimator) rates $r_i = s_i/t_i$. The cases in our tables are ordered by increasing raw failure rate, r_i.

Table 1. Air-Conditioner Failures (t_i in thousands of hours)

Aircraft no.	s_i	t_i	r_i
11	2	.623	3.21
9	9	1.800	5.00
5	14	1.832	7.64
4	15	1.819	8.25
12	12	1.297	9.25
10	6	.639	9.39
2	23	2.201	10.45
3	29	2.422	11.97
1	6	.493	12.17
13	16	1.312	12.20
7	27	2.074	13.02
8	24	1.539	15.59
6	30	1.788	16.78

Proschan (1963) and Ascher and Feingold (1978) noted that four of the air conditioners in the list were overhauled once during their observation periods. Ascher and Feingold concluded that there is evidence for changed reliability in three of the four cases experiencing overhaul. Our initial work analyzes the data without explicit recognition of such effects, which could be quite important in practice (see Table 1). A maximum likelihood ratio test (Cox and Lewis 1969, pp. 235–236) further indicates that the individual parameters are significantly different at about a 2% level. Thus these data can be expected to exhibit some between-rate variability as measured by a scale (e.g., standard deviation) parameter of the superpopulation.

2.2 Loss of Feedwater Flow

Table 2 presents a set of data referring to the rates of loss of feedwater flow for a collection of nuclear power generation systems (see Kaplan 1983). This class of "initiating events" is important in the probabilistic risk assessment (PRA) of nuclear plants. It was treated in an empirical Bayesian fashion by Kaplan, although he did not use that terminology.

Table 2. Loss of Feedwater Flow (t_i in years)

System$_i$	s_i	t_i	r_i
3	0	8	.0
19	0	2	.0
1	4	15	.27
7	2	5	.4
18	1	2	.5
8	4	4	1.0
16	3	3	1.0
25	1	1	1.0
4	10	8	1.3
10	4	3	1.3
15	4	3	1.3
5	14	6	2.3
13	10	4	2.5
27	5	2	2.5
20	3	1	3.0
9	13	4	3.3
2	40	12	3.3
26	10	3	3.3
12	14	4	3.5
14	7	2	3.5
24	12	3	4.0
28	16	4	4.0
29	14	3	4.7
21	5	1	5.0
30	58	11	5.3
17	11	2	5.5
22	6	1	6.0
6	31	5	6.2
11	27	4	6.8
23	35	5	7.0

2.3 Pump Reliability Data at a Pressurized Water Reactor Nuclear Power Plant

In Table 3, there appears a small set of data representing failures of pumps in several systems of the nuclear plant Farley 1. The apparent variation in failure rates has several possible sources; some are mentioned later. These data may be found in an Electric Power Research Institute (EPRI) report (Worledge, Stringham, and McClymont 1982).

3. SPECIFIC PEB MODELS

Consider two parametric families as representations of an assumed superpopulation for the event rates. These are (a) the centered and scaled log-Student-t, which includes the lognormal when the degrees of freedom parameter, n, becomes infinite (note that n is a tuning parameter, having nothing to do with a sample size) and (b) the gamma.

Form (a), the log-Student, is of potential interest in PRA of nuclear power systems in which the lognormal form has long been used to characterize variability between failure rates for various equipments (see Rasmussen et al. 1975; Kaplan 1983). The log-Student generalizes the lognormal setup, admitting systematically heavier-than-normal Gaussian tails and so allowing for a greater-than-Gaussian propensity for extreme outliers for the rates. The tail behavior is regulated by n, the Student degrees of freedom parameter. We do not here attempt to estimate n from data, but treat it as a tuning parameter, much in the manner of the tuning constant, c, appearing in biweight regression (see Mosteller and Tukey 1977). Form (b), the gamma, is the natural conjugate prior associated with the Poisson likelihood and hence yields pleasant analytical simplicity.

Here are the formal descriptions of the PEB models considered.

Log-Student. Stage 1 (sampling individual rates from the superpopulation):

$$\lambda_i = \exp(\varepsilon_i)$$

$$\varepsilon_i \sim g_\varepsilon(z; \mu, \tau; n) = \frac{C(n)}{[1 + ((z-\mu)/\tau)^2(1/n)]^{(n+1)/2}}, \tag{3.1}$$

where $C(n)$ is the appropriate normalizing constant and $\{\varepsilon_i, i = 1, 2, \ldots, I\}$ is a sequence of independent random variables.

Stage 2 (observations from the individual rates):

$$s_i \mid \lambda_i \sim \text{Poisson } (\lambda_i t_i). \tag{3.2}$$

Apparently $\varepsilon_i \sim \Phi((z - \mu)/\tau)$, the normal distribution, as $n \to \infty$, this being the lognormal model favored by many PRA analysts. In general,

$$\text{var}[\varepsilon_i] = \text{var}[\ln \lambda_i] = (n/(n-2))\tau^2, \qquad n > 2.$$

Gamma. Stage 1:

$$\lambda_i \sim g_\lambda(w; \alpha, \beta) = e^{-\alpha w}(\alpha(\alpha w)^{\beta-1}/\Gamma(\beta)). \tag{3.3}$$

Stage 2:

$$s_i \mid \lambda_i \sim \text{Poisson } (\lambda_i t_i). \tag{3.4}$$

There is no fundamental justification for either parametric superpopulation form. In general, the log-Student is appealing because of its controllable long

Table 3. Pump Failures (t_i in thousands of hours)

System$_i$	s_i	t_i	r_i
1	5	94.320	5.3×10^{-2}
2	1	15.720	6.4×10^{-2}
3	5	62.880	8.0×10^{-2}
4	14	125.760	11.1×10^{-2}
5	3	5.240	57.3×10^{-2}
6	19	31.440	60.4×10^{-2}
7	1	1.048	95.4×10^{-2}
8	1	1.048	95.4×10^{-2}
9	4	2.096	191×10^{-2}
10	22	10.480	209.90×10^{-2}

tails (the tails get long as n decreases) and the ease of interpretation of normal variation, whereas the gamma has mathematical convenience to recommend it. Neither represents truly eccentric behavior such as multimodality or extensive asymmetry—features that cannot be ruled out in real data. See Laird (1978) and Copas (1984) for nonparametric approaches to this problem and Tukey (1974) for an exploratory, totally nonprobabilistic approach to a large data set of similar structure.

4. FITTING THE SUPERPOPULATION MODELS: STAGE 1

Given data of the form exhibited in Tables 1, 2, and 3, it is possible to estimate the parameters in the superpopulation form by various methods. We examine two, simple moment matching and maximum likelihood.

4.1 Crude Moment Matching

From the Poisson assumption and familiar conditioning arguments, one can obtain these formulas:

$$E[\mathbf{s}_i \mid \boldsymbol{\lambda}_i] = \boldsymbol{\lambda}_i t_i = \text{var}[\mathbf{s}_i \mid \boldsymbol{\lambda}_i],$$
$$E[\mathbf{s}_i] = E[\boldsymbol{\lambda}_i]t_i,$$
$$\text{var}[\mathbf{s}_i] = E\{\text{var}[\mathbf{s}_i \mid \boldsymbol{\lambda}_i]\} + \text{var}\{E[\mathbf{s}_i \mid \boldsymbol{\lambda}_i]\}. \quad (4.1)$$

So, unconditionally,

$$E[\mathbf{s}_i] = E[\boldsymbol{\lambda}]t_i,$$
$$\text{var}[\mathbf{s}_i] = E[\boldsymbol{\lambda}]t_i + \text{var}[\boldsymbol{\lambda}]t_i^2. \quad (4.2)$$

Consequently if the raw rates are modeled by $\mathbf{r}_i = \mathbf{s}_i/t_i$,

$$E[\mathbf{r}_i] = E[\boldsymbol{\lambda}] \quad (4.3)$$

and

$$\text{var}[\mathbf{r}_i] = \text{var}[\boldsymbol{\lambda}] + E[\boldsymbol{\lambda}]\,(1/t_i), \quad (4.4)$$

which suggests that crude estimates for $E[\boldsymbol{\lambda}]$ and $\text{var}[\boldsymbol{\lambda}]$ can be obtained by matching moments:

$$\hat{E}[\boldsymbol{\lambda}] = \bar{r}, \qquad \hat{\text{var}}[\boldsymbol{\lambda}] = s_r^2 - \bar{r}\left(\frac{1}{I}\sum_{i=1}^{I} t_i^{-1}\right). \quad (4.5)$$

Now for the specific forms considered we know that for the lognormal,

$$E[\boldsymbol{\lambda}] = \exp(\mu + \tau^2/2); \quad \text{var}[\boldsymbol{\lambda}] = (E[\boldsymbol{\lambda}])^2(e^{\tau^2} - 1). \quad (4.6)$$

For the gamma,

$$E[\boldsymbol{\lambda}] = \beta/\alpha, \qquad \text{var}[\boldsymbol{\lambda}] = \beta/\alpha^2. \quad (4.7)$$

So both μ and τ^2, or α and β, can be assessed, perhaps inefficiently but very conveniently, by using (4.5) in conjunction with (4.6) or (4.7). Note that $E[\boldsymbol{\lambda}]$, and hence $\text{var}[\boldsymbol{\lambda}]$, is not finite for the log-Student model; therefore, this simplest moment matching procedure is inapplicable. Under the circumstance that s_i is large, accurate moment approximations for $\ln(\mathbf{r}_i) = \ln(\mathbf{s}_i/t_i)$ are obtainable for the Student superpopulation, provided that the Student parameter n is large enough (i.e., greater than 2). A more refined iterative estimating procedure for fitting the gamma was furnished by Hill, Heger, and Koen (1984), but the preceding formulas are extremely simple and useful for quick appraisals and handy as starts for iterative likelihood maximization.

4.2 Likelihood Methods

It is anticipated that the method of maximum likelihood will provide results superior to crude moment matching at the expense of greater computational effort, particularly for the log-Student form. Here are the likelihoods and comments concerning their maximization.

Log Student. The likelihood contribution of observation i is, up to irrelevant constants,

$$L_i = L_i(\mu, \tau; s_i, t_i; n) = \int_{-\infty}^{\infty} e^{-\lambda(z)t_i}[\lambda(z)]^{s_i} \times \frac{dz}{[1 + ((z-\mu)/\tau)^2(1/n)]^{(n+1)/2}}, \quad (4.8)$$

With $\lambda(z) = \exp(z)$, so the total likelihood is

$$L(\mu, \tau; \mathbf{s}, \mathbf{t}; n) = \prod_{i=1}^{I} L_i(\mu, \tau; \mathbf{s}, \mathbf{t}; n). \quad (4.9)$$

The integration and subsequent maximization must be carried out numerically. Integration has been performed by several alternative Gauss–Hermite procedures. The first begins by approximating the integral by Laplace's method and concludes by Gauss–Hermite integration of a correction term remaining after the Laplace effect is removed [see Gaver (1985) for details]; for brevity, this process will be called LGH. The second is a direct Gauss–Hermite procedure adapted from Naylor and Smith (1982) (we are grateful to J. C. Naylor for furnishing a FORTRAN program that has served as the basis for this aspect of our work); call this GH. The maximization was accomplished in the first method by a refined grid search and in the second by a quasi-Newton procedure adapted from IMSL subroutine ZXMIN, operating on the log-likelihood surface. The classical EM algorithm discussed by Dempster, Laird, and Rubin (1977) is applicable for estimating the superpopulation parameters, but it does not directly produce approximate confidence regions, as obtained hereafter.

Gamma. The likelihood contribution of observation i can be derived by integration and is the negative binomial expression

$$L_i(\alpha, \beta; s_i, t_i) = \frac{\Gamma(s_i + \beta)}{\Gamma(\beta)} \frac{t_i^{s_i} \alpha^\beta}{(t_i + \alpha_i)^{s_i+\beta}}. \quad (4.10)$$

In view of independence, a product of these contributions provides the total likelihood, as in (4.9). Maximization of the log-likelihood has then been carried out by the IMSL procedure. The work of Anscombe (1950) on asymptotic properties of moment and MLE estimates is somewhat relevant here.

The numerical results obtained from applying the preceding procedures to the three illustrative data sets are given in Figures 1, 2, and 3. To ease the comparison of the log-Student and gamma analyses, we have reparameterized the gamma in terms of μ and τ, the latter being the parameters of a lognormal. Thus the μ and τ lognormal values that match the first two gamma moments are

$$\mu = \ln[(\beta/\alpha)/\sqrt{1 + 1/\beta}], \qquad \tau = \sqrt{\ln(1 + 1/\beta)}; \quad (4.11)$$

these expressions have been used to parameterize the gamma likelihood for graphical display.

4.3 Approximate Confidence Regions for Superpopulation Parameters

The likelihood ratio test procedure has been used to define an approximate joint confidence region for μ and τ for the two superpopulation model families. The procedure specifies that all (μ, τ) values such that

$$-2[\ln(L(\mu, \tau; \mathbf{s}, \mathbf{t}; n)/L(\hat{\mu}, \hat{\tau}; n))] \le \chi_2^2(1 - \alpha),$$

where $(\hat{\mu}, \hat{\tau})$ is the MLE, constitute an approximate $(1 - \alpha) \cdot 100\%$ confidence region. The regions obtained for the three sets of data appear on the figures. The somewhat eccentric, but unimodal, shape of the log-likelihood surface is exhibited by the confidence contour plots; a bit more symmetry can in principle be obtained by reparameterizing in terms of $\ln \tau$, but for our data sets the effect was not dramatic. The confidence contours are roughly elliptical with the ellipse semi-axes nearly parallel to the $\mu - \tau$ axes; an analysis based on the simplifying assumption that $\hat{\mu}$ and $\hat{\tau}$ are independently bivariate normal actually works reasonably well for our data. The ellipticity tends to disappear when the data set is small and contains several $s_i = 0$ values; as anticipated the region then often intersects the $\tau = 0$ axis, suggesting that the data are consistent with a single underlying parameter value: $\lambda_i = \lambda$ $(i = 1, 2, \ldots, I)$ if the intersection is pronounced.

5. INDIVIDUALIZED ("SHRUNKEN" OR "POOLED") ESTIMATES

If the true values of μ and τ (log-Student) or α and β (gamma) superpopulations were available, then an obvious step would be to compute the Bayes posterior of $\boldsymbol{\varepsilon}_i = \ln \boldsymbol{\lambda}_i$ in the gamma case, given the value of $\mathbf{s}_i$. Then any point or interval estimates desired could be computed. Such calculations must be done numerically for the log-Student family, but they are eased in the gamma case by the conjugate prior assumption. If the values of μ and τ are estimated from data, as suggested here, then approximate superpopulations can be derived by replacing (μ, τ) by $(\hat{\mu}, \hat{\tau})$ and calculating the corresponding Bayes estimates [see Deely and Lindley (1981) for a discussion of this empirical Bayes approach]. Morris (1983) and Hill et al. (1984) suggested refinements to the simple procedure described. Use of the approximate individualized superpopulations (approximate Bayesian posteriors) then leads to point estimates and intervals. We have chosen to first calculate (a) the *means*, $\bar{\varepsilon}_i = E[\boldsymbol{\varepsilon}_i \mid \mathbf{s}_i]$, of the posterior distributions for the individual unit log rates, ε_i, in the illustrative data sets [these can be compared with ordinary log raw rates, $\ln(s_i/t_i)$]; (b) the *standard deviations*, $\sigma_i = [\mathrm{var}[\boldsymbol{\varepsilon}_i \mid \mathbf{s}_i]]^{1/2}$, of the posterior distributions; (c) the approximately 95% upper tolerance limits (or 95% one-sided Bayes credibility intervals) for each unit, based on a normal approximation: $\tilde{\varepsilon}_i(.95) = \bar{\varepsilon}_i + 1.645\sigma_i$; and (d) the upper confidence limits for the credibility intervals (c) that recognize sampling variability in $\hat{\mu}$ and $\hat{\tau}$. We are encouraged to believe that such intervals are reasonable by looking at plots of the posterior densities of the ε_i (see Figs. 1, 2, and 3). More exact calculations are possible by numerical integration of the posterior. Explicit expressions for the preceding quantities are as follows:

Log-Student. The approximate conditional means and second moments are cases of

$$E[\boldsymbol{\varepsilon}_i^k \mid s_i] = \frac{1}{L_i(\hat{\mu}, \hat{\tau}; s_i, t_i; n)} \int_{-\infty}^{\infty} z^k e^{-\lambda(z)t_i}[\lambda(z)]^{s_i} \times \frac{dz}{[1 + ((z - \hat{\mu}/\hat{\tau})^2(1/n)]^{(n+1)/2}}, \quad (5.1)$$

again integrated by Gauss–Hermite quadrature. The normalized integrand of (5.1), exclusive of z^k, is the approximate Bayes posterior density of $\boldsymbol{\varepsilon}_i$, given s_i.

Gamma. The mean and variance of the approximate gamma posterior have familiar pleasant explicit forms:

$$E[\boldsymbol{\lambda}_i \mid \mathbf{s}_i] = (\mathbf{s}_i + \hat{\beta})/(t_i + \hat{\alpha}) \quad (5.2)$$

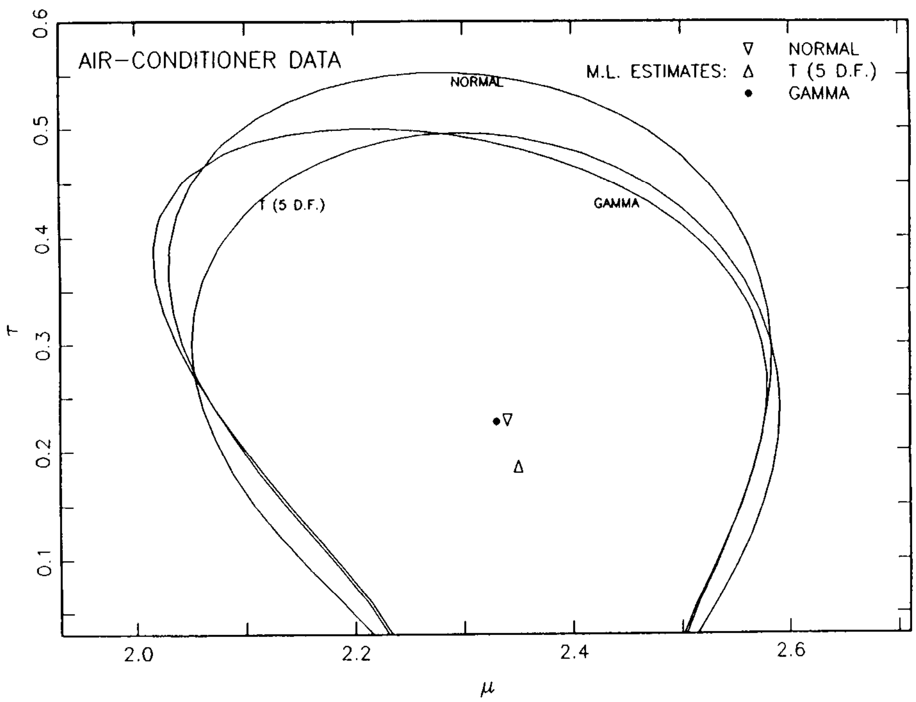

Figure 1. 95% Confidence Regions for μ, τ: Air-Conditioner Data.

and

$$\mathrm{var}[\lambda_i \mid \mathbf{s}_i] = (\mathbf{s}_i + \hat{\beta})/(t_i + \hat{\alpha})^2. \quad (5.3)$$

There are no such simple expressions for $\tilde{\varepsilon}_i = \ln \lambda_i$ in the gamma case, but the posterior moments have been computed by Gauss–Hermite quadrature applied to the log-transformed gamma density.

Analytical Approximations to the Estimates of Individual Rate

Although the preceding numerical computations are feasible, it is often useful to have relatively simple and explicit, if approximate, formulas for point estimates and posterior densities. One such can be

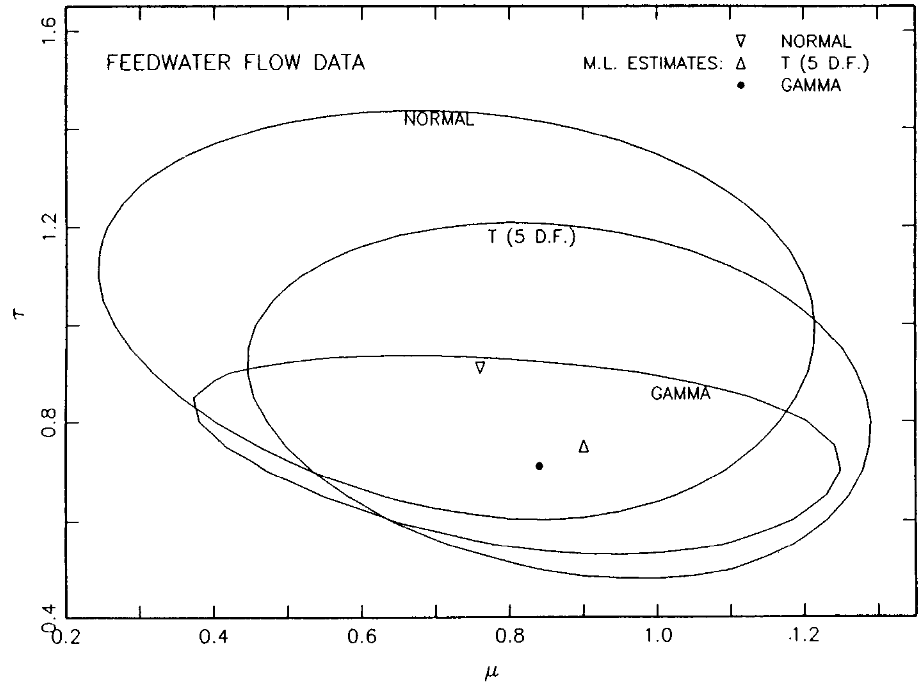

Figure 2. 95% Confidence Regions for μ, τ: Feedwater Flow Data.

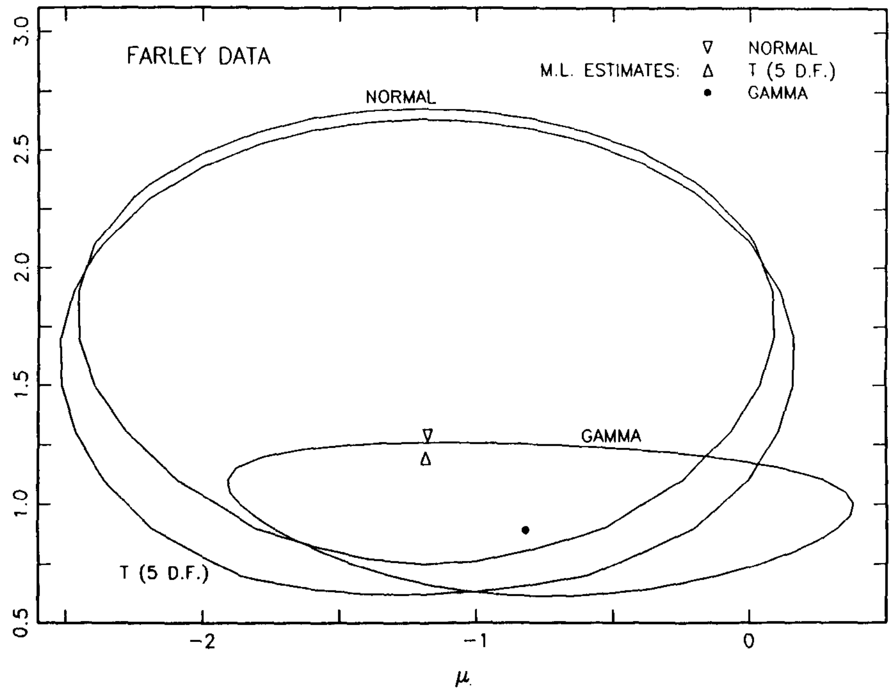

Figure 3. 95% Confidence Regions for μ, τ: Farley Data.

derived for the log-Student model by writing the posterior density as

$$g_\varepsilon(z \mid s_i) = Ke^{-(1/2)Q(z)} \tag{5.4}$$

and then approximating $Q(z)$ by a quadratic $q(z) = \frac{1}{2}([z - \mu(z)]/\sigma(z))^2$, in the manner of Laplace's method (see de Bruijn 1958; for applications to Bayesian statistics, see Mosteller and Wallace 1964, Tierney and Kadane 1986). Differentiation shows that the minimum of $Q(z)$ occurs at $\hat{\varepsilon}_i \simeq \bar{\varepsilon}_i$, where $\hat{\varepsilon}_i$ is the modal, or maximum likelihood, estimate of $\boldsymbol{\varepsilon}_i \mid \mathbf{s}_i$, and $\bar{\varepsilon}_i$ is the posterior mean.

Log-Student. The derivative of

$$Q(z) = -e^z t_i + s_i z - \left(\frac{n+1}{2}\right)\ln\left[1 + \left(\frac{z - \hat{\mu}}{\hat{\tau}}\right)^2 (1/n)\right] \tag{5.5}$$

set equal to 0 yields an estimating equation that can be written as follows:

$$\hat{\lambda}_i = e^{\hat{\varepsilon}_i} = \left[s_i - \left(\frac{\hat{\varepsilon}_i - \hat{\mu}}{\hat{\tau}^2}\right) w_n(\hat{\varepsilon}_i)\right](1/t_i), \tag{5.6}$$

where the *weight*

$$w_n(\hat{\varepsilon}_i) = \left(\frac{n+1}{n}\right)\frac{1}{1 + [(\hat{\varepsilon}_i - \hat{\mu})/\hat{\tau}]^2(1/n)}. \tag{5.7}$$

Graphical or analytical examination reveals the possibility of two solutions of $Q'(z) = 0$, one very near $\hat{\mu}$ and the other near $\ln(s_i/t_i)$, corresponding to a bimodal posterior. Convenient explicit analytical criteria for bimodality to occur are not available, but neither our present data sets, nor many others, have revealed such bimodality when the posterior was evaluated numerically and plotted. Our approach is to replace $w_n(\hat{\varepsilon}_i)$ by $\hat{w}_n = w_n(\ln(s_i t_i))$ and by $\hat{w}_n = w_n(\ln(1/(3t_i)))$, when $s_i = 0$. This approximate weight leads to unique solutions of (5.6) by Newton–Raphson iteration. An interesting and interpretable formula is obtained after one iteration, starting with $\hat{\varepsilon}_i(0) = \ln(s_i/t_i)$:

$$\begin{aligned}\hat{\varepsilon}_i(1) &= \frac{s_i \ln(s_i/t_i) + (\hat{\mu}/\hat{\tau}^2)\hat{w}_n}{s_i + (1/\hat{\tau}^2)\hat{w}_n} \\ &= \ln(s_i/t_i) - \frac{(\ln(s_i/t_i) - \hat{\mu})(1/\hat{\tau}^2)\hat{w}_n}{s_i + (1/\hat{\tau}^2)w_n}.\end{aligned} \tag{5.8}$$

This expression resembles the familiar Bayes normal-theory formula for combining prior mean and likelihood to obtain a linear estimator of the posterior mean. The difference is the presence of the weight $\hat{w}_n$, the effect of which is to reduce the influence of the mean of the superpopulation (prior) upon tail-discrepant observations. Discrepant observations ($\hat{w}_n$ small) are not heavily shrunken or pooled towards the estimated center, $\hat{\mu}$, whereas observations close to that center (w_n large) are shrunken in that direction. Actually, the net shrinkage is caused by the factor $[1/(\hat{\tau}^2)\hat{w}_n]/[s_i + (1/\hat{\tau}^2)\hat{w}_n]$, in which $\hat{w}_n$ plays an important but not exclusive role; the (approximate) variance s_i and $(\hat{\tau})^2$ are also significant [since $\hat{w}_n$ depends on $\hat{\tau}$ and $\ln(s_i/t_i)$, shrinkage is not linear]. Notice that as $n \to \infty$ and the log-Student approaches the lognormal, the discrepancy-tolerant effect diminishes; when $n = \infty$ there is always only one solution to (5.5) and shrinkage becomes linear (provided the effect of observation i on $\hat{\mu}$ and $\hat{\tau}$ is small, as it usually is). The variance of the posterior can be assessed from the second derivative of $Q(z)$; from (5.5),

$$\mathrm{var}[\varepsilon_i \mid s_i] = \sigma_i^2 \approx 1/[e^{\varepsilon_i} t_i + (1/\hat{\tau}^2)\hat{w}_n]. \tag{5.9}$$

This formula again exhibits the behavior of the variance associated with the posterior encountered when normal likelihoods are combined with normal conjugate priors, except that wildly tail-discrepant observations are substantially downweighted; automatically in such cases $\hat{\varepsilon}_i = \ln(s_i/t_i)$ and $\sigma_i^2 = 1/s_i$, as is essentially correct for a simple MLE if s_i is "large." In other words, our approximations (5.8) and (5.9) crudely treat $\ln(s_i/t_i)$ as normal with a conditional mean that is Student-t with nonnegligible variance. Such approximations are very convenient and lend themselves to simulation appraisals of the two-stage estimation procedure [see Gaver (1985) and a summary in Sec. 8 of this article].

Gamma. In the gamma case, it can be seen that

$$Q(z) = -\alpha' e^z + \beta' z, \tag{5.10}$$

where $\alpha' = t_i + \hat{\alpha}$ and $\beta' = s_i + \hat{\beta}$. Now differentiation again shows that

$$e^{\hat{\varepsilon}_i} = (\beta'/\hat{\alpha}) = (s_i + \hat{\beta})/(t_i + \hat{\alpha}) \tag{5.11}$$

and

$$\sigma_i^2 = 1/\beta' = 1/(s_i + \hat{\beta}). \tag{5.12}$$

Naturally, these formulas resemble the results for the log-Student superpopulation, but they contain no weight, $\hat{w}_n$, to reduce shrinkage effects upon tail-discrepant observations.

6. CONFIDENCE LIMITS

Since the estimates of posterior means, variances, and tolerance limits are functions of $\hat{\mu}$ and $\hat{\tau}$, it is desirable to place confidence limits on $\bar{\varepsilon}_i(\mu, \tau)$, $\sigma_i^2(\mu, \tau)$, and $\tilde{\varepsilon}_i(\mu, \tau)$. These may be based on the confidence contours of Figures 1–3 and are constructed numerically. We have supplied only upper 95% confidence contours, obtained by grid search over (μ, τ) space to maximize $\tilde{\varepsilon}_i(\mu, \tau)$, say, under the condition that (μ, τ) belongs to the appropriate confidence set; these limits are denoted by $\tilde{\tilde{\varepsilon}}_i$.

7. ANALYSIS OF DATA SETS

The estimation procedures described have been applied to the data sets of Tables 1–3. Complete tabulations are available from the authors; here we examine only those log parameter estimates that are at the low and high extremes for each data set and the middle or median level. Ordering of the rates is in terms of log raw rates. It is anticipated that the point estimates of the extreme individual rates will exhibit the greatest variation across estimation procedures (superpopulation models), whereas the middle values will be roughly in agreement. Owing to the partial pooling effect, the middle values will tend to exhibit somewhat smaller posterior variation than the extremes. The normal superpopulation model tends to shrink more extensively than the other superpopulation models. By and large, these effects are observed. Numerical results are summarized in

Table 4. Air-Conditioner Data Analysis

Ranked observations	Estimates					Intervals						
	$\hat{\varepsilon}(r)$	$\hat{\varepsilon}(1, 5)$	$\bar{\varepsilon}(5)$	$\bar{\varepsilon}(g)$	$\bar{\varepsilon}(\infty)$	$\tilde{\varepsilon}(r)$	$\tilde{\varepsilon}(1, 5)$	$\tilde{\varepsilon}(5)$	$\tilde{\tilde{\varepsilon}}(5)$	$\tilde{\varepsilon}(g)$	$\tilde{\tilde{\varepsilon}}(g)$	$\tilde{\varepsilon}(\infty)$
Smallest ($i = 1$)	1.17 (.71)	1.94 (.35) [.13]	2.14 (.24)	2.13 (.22)	2.16 (.20)	2.33 [2.03]	2.51	2.54	2.73	2.50	2.61	2.49
Median ($i = 7$)	2.35 (.21)	2.34 (.13) [1.20]	2.34 (.14)	2.34 (.16)	2.34 (.15)	2.70 [2.66]	2.56	2.58	2.67	2.59	2.67	2.59
Largest ($i = 13$)	2.82 (.18)	2.66 (.16) [.52]	2.60 (.17)	2.60 (.14)	2.61 (.17)	3.12 [3.04]	2.92	2.88	3.05	2.84	3.03	2.86

Superpopulation parameters
Stud (5): For $\hat{\mu}$—MLE, 2.35; MM, —; for $\hat{\tau}$—MLE, .19; MM, —.
Normal: For $\hat{\mu}$—MLE, 2.34; MM, 2.31; for $\hat{\tau}$—MLE, .23; MM, .24.
Gamma: For $\hat{\mu}$—MLE, 2.33; MM, —; for $\hat{\tau}$—MLE, .23; MM, — ($\hat{\alpha} = 1.74$, $\hat{\beta} = 18.41$).

NOTE: MLE indicates maximum likelihood estimator; MM indicates moment method.

Table 5. Loss of Feedwater Flow Data Analysis

Ranked observations	*Estimates*					*Intervals*						
	$\hat{\varepsilon}(r)$	$\hat{\varepsilon}(1, 5)$	$\bar{\varepsilon}(5)$	$\bar{\varepsilon}(g)$	$\bar{\varepsilon}(\infty)$	$\tilde{\varepsilon}(r)$	$\tilde{\varepsilon}(1, 5)$	$\tilde{\varepsilon}(5)$	$\tilde{\tilde{\varepsilon}}(5)$	$\tilde{\varepsilon}(g)$	$\tilde{\tilde{\varepsilon}}(g)$	$\tilde{\varepsilon}(\infty)$
Smallest (i = 1)	(−3.18) (.57)	−1.74 (.76) [.17]	−2.20 (1.33)	−2.08 (.96)	−1.30 (.53)	−2.24 [−]	−.49	−.01	.10	−.51	.03	−.42
Smallest nonzero (i = 3)	−1.32 (.50)	−.96 (.39)	−1.05 (.46)	−1.13 (.45)	−.97 (.38)	−.50 [−.66]	−.50	−.32	−.25	−.39	−.09	−.35
Medians (i = 15)	1.10 (.57)	1.02 (.45) [1.18]	.95 (.47)	.98 (.50)	.91 (.51)	2.04 [1.58]	1.76	1.73	1.87	1.80	1.82	1.75
(i = 16)	1.18 (.28)	1.14 (.26) [1.17]	1.11 (.27)	1.13 (.27)	1.11 (.27)	1.641 [1.58]	1.64	1.55	1.59	1.57	1.58	1.56
Largest (i = 30)	1.95 (.17)	1.90 (.17) [.86]	1.89 (.17)	1.88 (.17)	1.89 (.17)	2.23 [2.20]	2.23	2.17	2.19	2.15	2.18	2.17

Superpopulation parameters
Stud (5): For $\hat{\mu}$—MLE, .76; MM, —; for $\hat{\tau}$—MLE, .91; MM, —.
Normal: For $\hat{\mu}$—MLE, .93; MM, .94; for $\hat{\tau}$—MLE, .72; MM, .57.
Gamma: For $\hat{\mu}$—MLE, .83; MM, —; for $\hat{\tau}$—MLE, .71; MM, — ($\hat{\alpha}$ = .52, $\hat{\beta}$ = 1.53).

NOTE: MLE indicates maximum likelihood estimator; MM indicates moment method.

Tables 4, 5, and 6. For a visual notion of the posterior densities from our data sets, see Figures 4–6.

7.1 Table Notation

The following notation has been used under the *Estimates* column headings:

1. $\hat{\varepsilon}(r) = \ln(s_i/t_i) = \ln(r_i)$.
2. $\hat{\varepsilon}(1, n)$ is the linearized posterior mode, Student (n) prior, ($n = 5$) here; see (5.8).
3. $\bar{\varepsilon}(n)$ is the posterior mean, Student (n) prior.
4. $\bar{\varepsilon}(g)$ is the posterior mean, Gamma prior.
5. $\bar{\varepsilon}(\infty)$ is the posterior mean, normal/Gauss prior.

The numbers in parentheses under each of the preceding notations in Tables 4, 5, and 6 are the stan-

Table 6. Pump Data Analysis

Ranked observations	*Estimates*					*Intervals*						
	$\hat{\varepsilon}(r)$	$\hat{\varepsilon}(1, 5)$	$\bar{\varepsilon}(5)$	$\bar{\varepsilon}(g)$	$\bar{\varepsilon}(\infty)$	$\tilde{\varepsilon}(r)$	$\tilde{\varepsilon}(1, 5)$	$\tilde{\varepsilon}(5)$	$\tilde{\tilde{\varepsilon}}(5)$	$\tilde{\varepsilon}(g)$	$\tilde{\tilde{\varepsilon}}(g)$	$\tilde{\varepsilon}(\infty)$
Smallest (i = 1)	−2.94 (.45)	−2.75 (.39) [.84]	−2.84 (.42)	+2.89 (.43)	−2.83 (.40)	−2.20 [−2.33]	−2.10	−2.16	−2.04	−2.17	−2.08	−2.17
Medians (i = 5)	−.56 (.58)	−.69 (.54) [1.14]	−.81 (.56)	−.67 (.55)	−.79 (.57)	.39 [.18]	.20	.12	.29	.23	.44	.15
(i = 6)	−.50 (.23)	−.53 (.23) [1.13]	−.56 (.23)	−.55 (.23)	−.57 (.23)	−.12 [−.16]	.−15	−.18	−.14	−.15	−.10	−.17
Largest (i = 10)	.74 (.21)	.69 (.22) [.79]	.67 (.22)	.64 (.21)	.67 (.22)	1.08 [1.06]	1.05	1.03	1.06	.99	1.08	1.03

Superpopulation parameters
Stud (5): For $\hat{\mu}$—MLE, −1.18; MM, —; for $\hat{\tau}$—MLE, 1.29; MM, —.
Normal: For $\hat{\mu}$—MLE, −1.19; MM, −.55; for $\hat{\tau}$—MLE, 1.19; MM, .71.
Gamma: For $\hat{\mu}$—MLE, −.83; MM, —; for $\hat{\tau}$—MLE, .89; MM, — ($\hat{\alpha}$ = 1.27, $\hat{\beta}$ = .82).

NOTE: MLE indicates maximum likelihood estimator; MM indicates moment method.

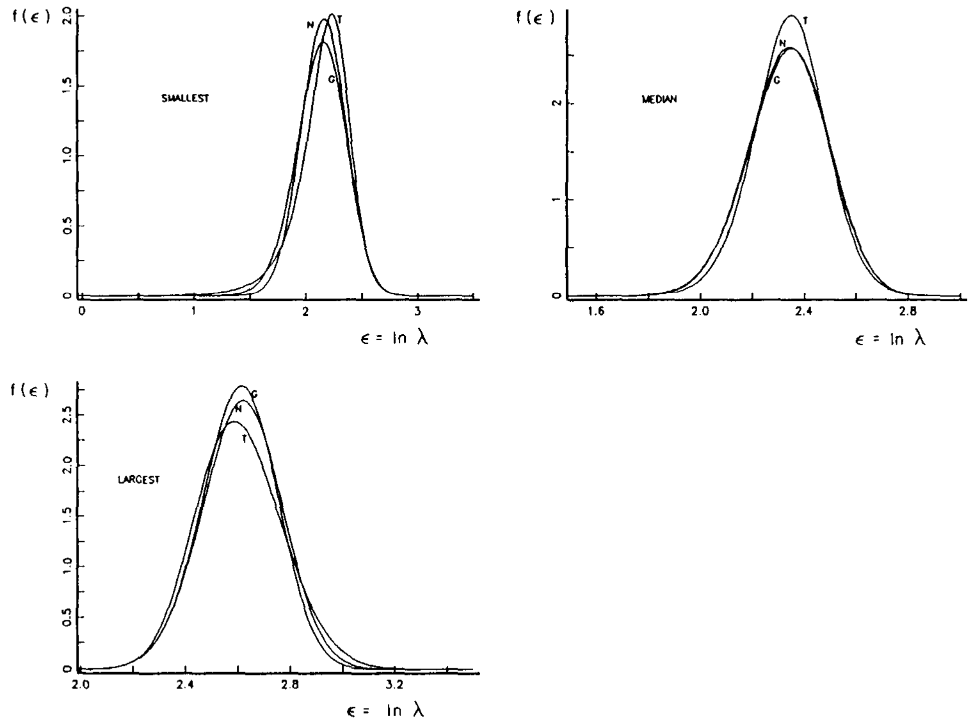

Figure 4. Posterior Densities: Air-Conditioner Data.

dard deviations of the associated posteriors; posterior means and standard deviations are computed either by numerical integration, in cases 3 and 5, or by simple explicit approximate formulas in cases 1, 2, and 4. Under the *Intervals* headings, there are included approximate upper 95% Bayes credibility intervals based on a normal approximation [mean + (1.645) (standard deviation)]; these are the

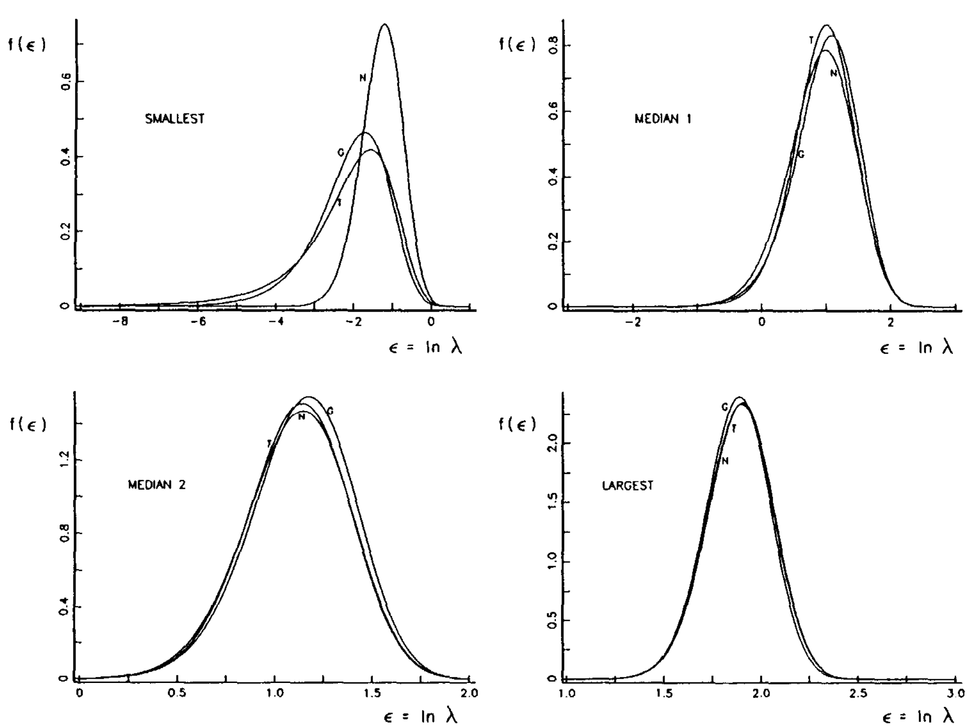

Figure 5. Posterior Densities: Feedwater Flow Data.

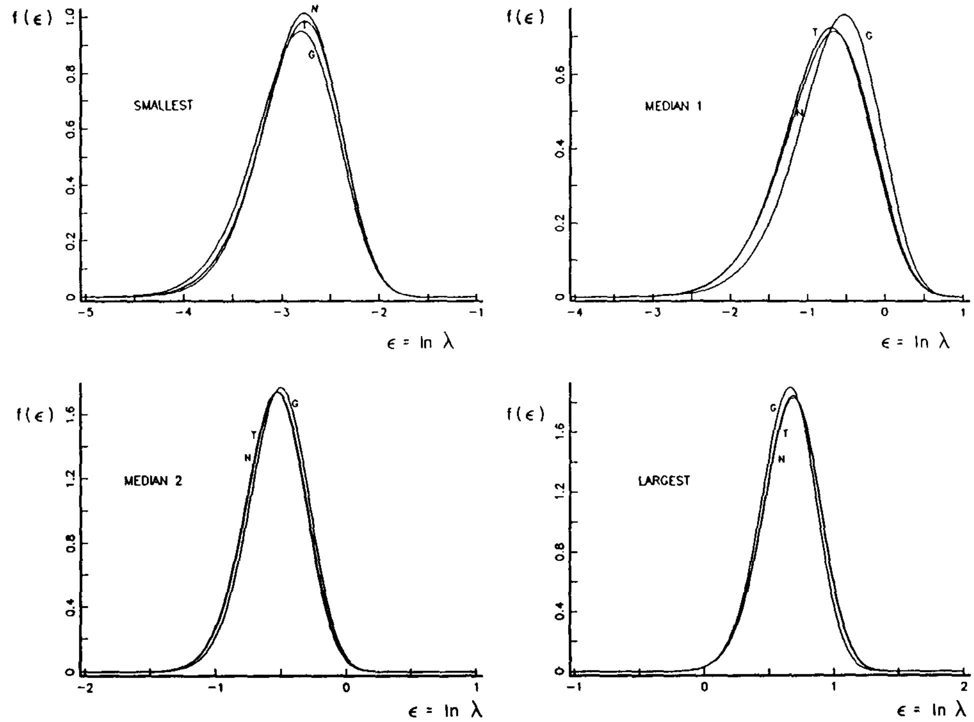

Figure 6. Posterior Densities: Farley Data.

following.

6. $\tilde{\varepsilon}(r) = \hat{\varepsilon}(r) + 1.645\hat{\sigma}(r) = \hat{\varepsilon}_l(r) + 1.645\sqrt{1/s_i}$, using $1/s_i$, the simplest delta-method approximation to $\mathrm{var}[\ln(s_i/t_i)]$; in [ ] we quote the upper limit computed making use of the chi-squared distribution of the time to accumulate s failures, an approximation to the former.

7. $\tilde{\varepsilon}(1, n)$ is the same as the preceding, using the Student (n) prior ($n = 5$) and linearized estimates [see (5.8) and (5.9)].

8. $\tilde{\varepsilon}(n)$ is the same as 6, but using 3 and associated standard deviation.

9. $\tilde{\tilde{\varepsilon}}(n)$ is the upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as described in Section 6.

10. $\tilde{\varepsilon}(g)$ is the same as 6, using moments of log-gamma computed numerically.

11. $\tilde{\tilde{\varepsilon}}(g)$ is the upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as in Section 6.

12. $\tilde{\varepsilon}(\infty)$ is the same as 8, using estimated normal prior.

13. $\tilde{\tilde{\varepsilon}}(\infty)$ is the upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as in Section 6.

7.2 Air-Conditioner Data

Upward shrinkage of the smallest estimate, $\hat{\varepsilon}(r)$, is most pronounced for $\tilde{\varepsilon}(\infty)$, the normal prior, less so for the gamma, and still less so for the Student (5)'s; the simple linear approximation least so. The linearized Student (5) procedure gives a small weight to the smallest observed rate. Upper interval boundaries differ less than point estimates, with the $\tilde{\tilde{\varepsilon}}$ levels only slightly above $\tilde{\varepsilon}$.

The middle estimate is shrunk not at all numerically by any of the point estimates, but the standard deviations of all shrunken/pooled estimates are about 70% of that of the raw estimate $\hat{\varepsilon}(r)$. Upper interval levels are correspondingly reduced.

The largest estimate is shrunk downward slightly and consistently by all estimates. Shrinkage is less extensive for the largest than the smallest; this can be partly explained by the weights: .52 versus .13.

7.3 Feedwater Flow Data

The smallest observation is a 0, and the crudely imputed rate is $\hat{\varepsilon}(r) = \ln(1/3t_i)$; it is enclosed in parentheses to signify its arbitrary nature. Here all point estimates provide some upward shrinkage, the normal prior estimate, $\tilde{\varepsilon}(\infty)$, shrinking upwards the most extensively; it also exhibits the smallest standard deviation. Here the Student (5) credibility and confidence intervals tend to be lower than the corresponding gamma and normal intervals.

The first middle estimate, $(i) = 15$ in rank, is shrunk downward by perhaps 10%; the most extensive shrinkage occurs for the normal model, $\tilde{\varepsilon}(\infty)$. Its shrunken standard deviation is about 80% of that of the raw for the Student model. Note that this observation involves $s = 3$ events over exposure time $t = 1$, a short history. By contrast, its neighboring middle value, $(i) = 16$ in rank, with the more extensive history $s = 13$ over $t = 4$, exhibits one-half the shrinkage and very little standard deviation decrease.

The largest estimate is shrunken nearly the same by all estimates; the upper intervals agree well internally, tending to be slightly below the interval raw rate interval, $\hat{\varepsilon}(r)$.

7.4 Farley Pump Data

The smallest observation, $\hat{\varepsilon}(r)$ in this data set, is shrunk toward the mean to essentially the same degree by each alternative point estimate; slightly less shrinkage occurs for the gamma and Student (5) models. The upper 95% credibility limits, $\bar{\varepsilon}$, also agree for all models, with $\hat{\varepsilon}(1, 5)$ being marginally the greatest. The upper confidence limits, $\tilde{\varepsilon}$, are in agreement as well.

The two median values at $(i) = 5, 6$ are individually treated consistently so far as point estimates go: All shrunken estimates reduce the log raw rate toward the mean, with the greater shrinkage occurring for $(i) = 5$ because of its smaller experience (failure count and exposure time). For the same reason, posterior standard deviations for $(i) = 5$ are more than twice as great as those for $(i) = 6$, and upper 95% credibility intervals and confidence limits reflect this difference as well.

The point estimates associated with the largest log raw rate all substantially agree in their modest downward shrinkage and the upper credibility and confidence limits. Again, the close agreement is attributable to the relatively extensive experience embodied in unit $(i) = 10$.

It is, however, worth notice that the estimated scale parameter, τ, is quite large for this data set. A plausible reason is that the units in the set are not truly homogeneous and that much of the large variability is explainable by classification or regression. Our estimation procedures tend to reflect this: Although weights $\hat{w}_n$ are rather similar for extreme and middle observations, the actual shrinkages are small even when there is little experience [e.g., for $(i) = 5$]. In fact, investigation reveals that the four pumps with the greatest experience (relatively large s and long t) all operate continuously, whereas the remaining six operate intermittently or on standby; these latter display consistently higher failure rates than the former, so a dummy variable (continuous vs. intermittent) regression model should tend to reduce $\hat{\tau}$. In Figure 7 we exhibit the result of a reanalysis in which the two groups' estimates of μ and τ are found separately; the two point estimate vectors are now much more consistent and the confidence regions are smaller and only partially overlap.

8. SIMULATION RESULTS

Limited simulation experiments have been carried out to evaluate some of the estimation procedures described. Here is the design [see Gaver (1985) for further details].

First, the superpopulation form for the Poisson log rates ε_i was taken for convenience to be a member of the controllably long-tailed Tukey h family: $\boldsymbol{\varepsilon}_i \sim \mu + \tau\mathbf{z}_i \exp(h\mathbf{z}_i^2)$ with $\mathbf{z}_i \sim N(0, 1)$ and h, the tail-stretching parameter, nonnegative (here $h = .15$) [see Hoaglin (1983) and Gaver (1983) for details concerning this family]. We wished to compare the treatment of the different rate values in a random sample from the superpopulation by various estimators, so ordered λ values were next created (and stored): $\lambda_{(i)} = \exp(\varepsilon_{(i)})$, $\varepsilon_{(i)} = \mu + \tau z_{(i)} \exp[hz_{(i)}^2]$, $z_{(i)}$ being the ith largest order statistic in a sample of size I from $N(0, 1)$. For $h > 0$ the extremes $\lambda_{(1)}$ and $\lambda_{(I)}$ tend to be outliers, and the median, $\lambda_{((I+1)/2)}$, is characteristic of a central value. Second, the $\lambda_{(i)}$ values were used to generate Poisson counts, $s_{(i)}$. Then Stage 1 and Stage 2 estimation processes were applied to estimate first μ, τ, and then the individual λ_i values. The speedy LGH procedure was used to estimate μ and τ, and these values were then used in conjunction with the approximation $\lambda_i = \exp(\varepsilon_i)$ that solves (5.6) by iteration. Detailed numerical quadrature using the GH procedure is perhaps superior, but it would have consumed more computer time. The squared differences of the estimated $\boldsymbol{\lambda}_i$ values and their true counterparts were then averaged over $S(=200)$ simulations and quoted as mean-squared errors (MSE's). Table 7 summarizes results for several such experiments. We have quoted the ordinary units estimate results as MLE, the results of applying the estimating formulas (5.6) as RS (restricted shrinkage, as governed by $\hat{w}_n$), and the results of applying (5.6) without the weight as SS (simple shrinkage); the latter represents approximately the effect of applying a log-Student model when $n = 50$.

The results obtained are suggestive if not dramatic. First, estimates of the superpopulation mean, μ, are nearly unbiased, whereas those for τ^2 appear biased low. Standard errors of estimates (the figures in parentheses) are, not surprisingly, substantial; apparently more simulation repetitions would be desirable. Nevertheless, comparison of the MSE figures for the various estimators implies that RS$(n = 4)$ has virtue: for the smallest and largest rates, $\lambda_{(1)}$ and $\lambda_{(15)}$, RS estimates resemble MLE performance, whereas SS

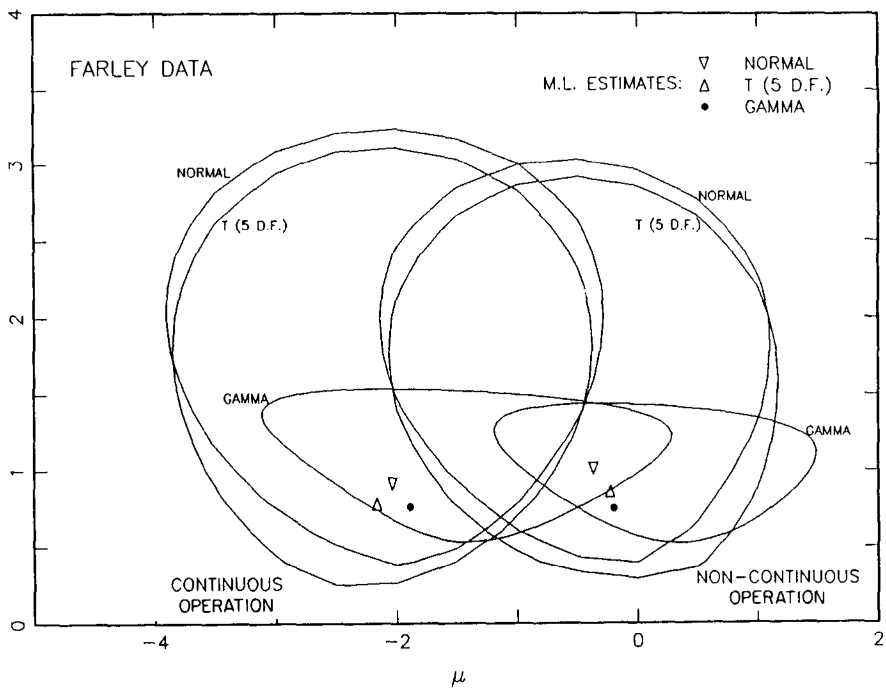

Figure 7. 95% Confidence Regions for μ, τ: Farley Data.

over-shrinks, for the middle value, $\lambda_{(8)}$, RS is far superior to the simple individual, unpooled MLE. The numerical differences in MSE shown in the table are small but real because of positive correlation between estimate values on each simulation experiment.

9. SUMMARY AND CONCLUSIONS

This article displays the results of analyzing several small batches (optimistically, but not realistically, random samples) of event rate data as if (a) parameters of each batch were drawn independently from a fixed superpopulation and then (b) the batches themselves were samples from random processes, here stationary Poisson or iid exponential-interval. Such is at least a pleasant fiction to be used as a starting point. Computational methods have been used to obtain estimates of superpopulation parameters, and from these pooled or shrunken individualized (log) rate estimates were obtained. Such PEB analyses were described by Hill et al. (1984), Deely and Lindley (1981), Hinde (1982), Kaplan (1983), and no doubt others. We extend these by introducing a heavy-tailed superpopulation form, the log-Student t, that allows for outliers or tail discrepancies incompatible with the lognormal/Gauss description. We call this the RPEB setup. The qualitative effect of such a generalization is revealed by appearance of the weight, $\hat{w}_n$, that selectively reduces the linear shrinkage towards a center [see (5.8)]. Thus $\hat{w}_n$ plays a role similar to that of an influence function in robust location estimation (see Mosteller and Tukey 1977, p. 351), although in the estimation of a single log rate it curbs the influence of the overall mean, $\hat{\mu}$, on that estimate if the data give evidence of extreme discrepancy. A more complete indication of the effect of an observation—that is, $\ln(s_i/t_i) = \hat{\varepsilon}_i(r)$—on its own shrinkage is given by the quantity $[1/\hat{\tau}^2]\hat{w}_n/[s_i + (1/\hat{\tau}^2)\hat{w}_n]$ appearing in the final expression in (5.8); both within variability, measured by $s_i(\approx \mathrm{var}[\hat{\varepsilon}_i(r)])$, and between variability, assessed by $\hat{\tau}^2$, play their parts along with $\hat{w}_n$. Besides providing insight, expressions like (5.8) and (5.9) seem to agree reasonably

Table 7. Selected Mean Squared Error Comparisons and Estimated Superpopulation Parameters

True values	Estimated	Estimator	$\lambda_{(1)}$ (small)	$\lambda_{(8)}$ (median)	$\lambda_{(15)}$ (large)
	(n = 4) $\hat{\mu}$ = −.97(.41) $\hat{\tau}^2$ = .17(.15)	RS	.016	.019	.33
μ = −1.0 τ^2 = .25		MLE	.007	.030	.32
	(n = 50) $\hat{\mu}$ = −.98(.45) $\hat{\tau}^2$ = .18(.15)	SS	.019	.020	.35
	(n = 4) $\hat{\mu}$ = −1.93(.50) $\hat{\tau}^2$ = .18(.17)	RS	.050	.0060	.28
μ = −2.0 τ^2 = .25		MLE	.0026	.014	.27
	(n = 50) $\hat{\mu}$ = −1.93(.52) $\hat{\tau}^2$ = .20(.18)	SS	.0054	.0057	.30

NOTE: J is 15 and h is .15 with 200 simulations. The Student df (tuning constant) n = 4, 50.

well with more exact results and are easy to compute, especially if one settles for inefficient moment estimators of superpopulation parameters.

As the tables and figures reveal, the example data analyses performed do not show enormous differences between lognormal, log-Student(5), and gamma superpopulation (Bayes prior) specifications, especially for the median and also for the largest batch values. The smallest batch observations are treated similarly by gamma and Student(5), with the normal/Gauss representation tending to shrink a "small" (zero) value more extensively than do the others up toward the center, μ, on the log scale. As anticipated, other analyses indicate even less tail shrinkage by Student(n) for $n < 5$. Estimation of n from the batch values would be of interest but is unlikely to be done with much accuracy from scanty data. This suggests that use of a gamma form for the prior and hence for the posterior may sometimes be relatively harmless. There is little evidence in our examples that overshrinkage of the largest values in a data set occurs when the gamma specification is used (although a small-n Student analysis could be performed as an indication of the possible extent of overshrinkage). Certainly the gamma is technically convenient for computing predictive estimates of reliabilities or availabilities of complex systems, integrations can often be carried out explicitly as Laplace transforms.

Of considerable interest would be the reduction of the apparent between-variability by classification or regression, as briefly illustrated for the Farley data. Research in this area is currently in progress, with promising results. If part of the between variability could be suitably accounted for, then estimators could be constructed that legitimately pool toward appropriate individualized centers, μ_i rather than μ, and outliers could be explained and reduced in effect. All of the preceding requires attention to collection of representative current data and the monitoring of analytical results over time to check for changes—for example, in basic parameters. Our present analysis is only a step in this direction. Further generalizations include analyses for failure-on-demand data, for which responses are binary and explanatory variables could include the time durations between inspections or serious activations—that is, those in response to a system demand. Analyses of complex redundant systems were proposed by Gaver and Lehoczky (1985).

ACKNOWLEDGMENTS

D. P. Gaver gratefully acknowledges research support from the Nuclear Safety and Analysis Department of the Electric Power Research Institute (EPRI) in Palo Alto, California, and from the Office of Naval Research, I. G. O'Muircheartaigh carried out this work while a visiting faculty member at the Naval Postgraduate School in Monterey, California. The advice and encouragement of David Worledge of EPRI has been most useful and welcome.

[Received May 1985. Revised July 1986.]

REFERENCES

Albert, J. H. (1979), "Robust Bayes Estimation," Mimeo Series 79-9, Purdue University, Dept. of Statistics.

Anscombe, F. J. (1950), "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions," *Biometrika*, 37, 358–382.

Ascher, A., and Feingold, H. (1978), "The Aircraft Air Conditioner Data Revisited," in *Proceedings of the Annual Reliability and Maintainability Symposium*, New York: Institute of Electrical and Electronic Engineers, pp. 153–159.

Berger, J. O. (1980), *Statistical Decision Theory*, New York: Springer-Verlag.

——— (1984), "The Robust Bayesian Viewpoint," in *Robustness of Bayesian Analyses*, ed. J. B. Kadane, Amsterdam: Elsevier.

Burridge, J. (1984), "Empirical Bayes Analysis of Survival Time Data," *Journal of the Royal Statistical Society*, Ser. B, 43, 65–76.

Copas, J. B. (1984), "A Computer-Intensive Non-Parametric Estimate of a Mixing Distribution, With Application to Poisson Failure Data," technical report, Naval Postgraduate School, Operations Research Dept.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, London: Methuen.

Cox, D. R., and Snell, E. J. (1980), *Applied Statistics*, London: Chapman & Hall.

De Bruijn, N. G. (1958), *Asymptotic Methods in Analysis*, New York: Interscience.

Deely, J. J., and Lindley, D. V. (1981), "Bayes Empirical Bayes," *Journal of the American Statistical Association*, 76, 833–841.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Gaver, D. P. (1963), "Random Hazard in Reliability Problems," *Technometrics*, 5, 211–226.

——— (1983), "Stochastic Modeling: Ideas and Techniques," in *Probability Theory and Computer Science*, eds. G. Louchard and G. Latouche, New York: Academic Press, pp. 3–49.

——— (1985), "Discrepancy Tolerant Hierarchical Poisson Event Rate Analyses," technical report, Naval Postgraduate School, Operations Research Dept.

Gaver, D., and Lehoczky, J. (1985), "Random Parameter Markov Population Process Models and Their Likelihood, Bayes, and Empirical Bayes Analysis," technical report, Naval Postgraduate School, Operations Research Dept.

Hill, J. R., Heger, A. S., and Koen, B. V. (1984), "The Application of Stein and Related Parametric Empirical Bayes Estimators to the Nuclear Plant Reliability Data Systems," Technical Report NUREG/CR-3637, University of Texas.

Hinde, J. (1982), "Compound Poisson Regression Models," in *GLIM 82: Proceedings of the International Conference in Generalized Linear Models*, ed. R. Gilchnist, Berlin: Springer-Verlag, pp. 109–121.

Hoaglin, G. C. (1983), "*g* and *h* Distributions," in *Encyclopedia of Statistics* (Vol. 3), New York: John Wiley.

Johnstone, I. (1984), "Admissibility, Difference Equations and Recurrence in Estimating a Poisson Mean," *The Annals of Statistics*, 12, 1173–1198.

Kaplan, S. (1983), "On a Two-stage Bayesian Procedure for Determining Failure Rates From Experimental Data," *IEEE Transactions on Power Apparatus and Systems*, 102, 195–202.

Laird, N. (1978), "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73, 805–811.

Martz, H. F., and Waller, R. A. (1982), *Bayesian Reliability Analysis*, New York: John Wiley.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman & Hall.

Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications" (with discussion), *Journal of the American Statistical Association*, 78, 47–65.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.

Mosteller, F., and Wallace, D. L. (1964), *Inference and Disputed Authorship: The Federalist*, Reading, MA: Addison-Wesley.

Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics*, 31, 214–225.

Proschan, F. (1963), "Theoretical Explanation of Observed Decreasing Failure Rate," *Technometrics*, 5, 375–383.

Rasmussen, N., et al. (1975), "Reactor Safety Study: An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants," Report NUREG-75/014, WASH 1400, Washington, DC: Nuclear Regulatory Commission.

Reynolds, D. S., and Savage, I. R. (1971), "Random Wear Models in Reliability Theory," *Advances in Applied Probability*, 3, 229–248.

Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Tukey, J. W. (1974), "Named and Faceless Values: An Initial Exploration in Memory of Prasanta C. Mahalanobis," *Sankhya*, Ser. A, 36, 125–176.

Worledge, D. H., Strigham, R. S., and McClymont, A. S. (1982), "PWR Power Plant Reliability Data," Interim Report NP-2592, Electric Power Research Institute.